(12)

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY
and
CENTER FOR BIOLOGICAL INFORMATION PROCESSING
WHITAKER COLLEGE

A.I. Memo No. 1348 January 1991
C.B.I.P Memo No. 63

Task and object learning in visual recognition

Shimon Edelman Heinrich H. Bülthoff Erik Sklar

Abstract

Human performance in object recognition changes with practice, even in the absence of feedback to the subject. The nature of the change can reveal important properties of the process of recognition. We report an experiment designed to distinguish between non-specific task learning and object-specific practice effects. The results of the experiment support the notion that learning through modification of object representations can be separated from secondary effects of practice, if appropriate response measures (specifically, the coefficient of variation of response time over views of an object) are used. The present results, obtained with computer-generated amoeba-like objects, corroborate previous findings regarding the development of canonical views and related phenomena with practice.

This report describes research done at the Massachusetts Institute of Technology within the Center for Biological Information Processing in the Department of Brain and Cognitive Sciences and Whitaker College. The Center's research is sponsored by grant N00014-91-J-1270 from the Office of Naval Research (ONR), Cognitive and Neural Sciences Division; and by National Science Foundation grant IRI-8719394. SE is at the Department of Applied Mathematics and Computer Science, The Weizmann Institute of Science, Rehovot 76100, Israel. HHB and ES are at the Department of Cognitive and Linguistic Sciences, Brown University, Providence, RI 02192.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	January 1992	memorandum

4. TITLE AND SUBTITLE

Task and Object Learning in Visual Recognition

5. FUNDING NUMBERS

IRI-8719394
N00014-91-J-1270

6. AUTHOR(S)

Shimon Edelman, Heinrich H. Bülthoff, and Erik Sklar

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Artificial Intelligence Laboratory
545 Technology Square
Cambridge, Massachusetts 02139

8. PERFORMING ORGANIZATION REPORT NUMBER

AIM 1348
C.B.I.P. 63

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research
Information Systems
Arlington, Virginia 22217

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

None

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Distribution of this document is unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Human performance in object recognition changes with practice, even in the absence of feedback to the subject. The nature of the change can reveal important properties of the process of recognition. We report an experiment designed to distinguish between non-specific task learning and object-specific practice effects. The results of the experiment support the notion that learning through modification of object representations can be separated from secondary effects of practice, if appropriate response measures (specifically, the coefficient of variation of response time over views of an object) are used. The present results, obtained with computer-generated amoeba-like objects, corroborate previous findings regarding the development of canonical views and related phenomena with practice.

14. SUBJECT TERMS (key words)	15. NUMBER OF PAGES
visual recognition learning	12
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED

1 Introduction

1.1 General background

Recent results in visual psychophysics indicate that practice, including mere repeated exposure without feedback, affects human performance in object recognition in several ways (Tarr and Pinker, 1989; Edelman et al., 1989; Edelman and Weinshall, 1991). In a typical recognition experiment, the subject is first shown the target object, possibly from several viewpoints and/or in motion. The subject is then asked to recognize various views, either familiar or unfamiliar, of objects that may themselves be novel. The present discussion is limited to the case of familiar test views (see (Rock and DiVita, 1987; Edelman and Bülthoff, 1990; Bülthoff and Edelman, 1992) for a treatment of the question of generalization of recognition to novel views). We also assume that the task calls for subordinate-level recognition (that is, both the target and the non-target objects belong to the same basic category (Rosch et al., 1976); for basic-level recognition, see, e.g., (Biederman, 1987)).

Not unexpectedly, under these conditions practice causes a general reduction in mean response time (RT). In addition, if the subject is given feedback about the correctness of the response, the error rate (ER) undergoes a similar evolution.[1] However, it appears that practice also precipitates other changes in RT and ER. To understand the nature and possible cause of these changes, we must first consider two major characteristics of human performance in the recognition of previously seen views of 3D objects.

1.2 Canonical views and mental rotation in recognition

Two basic characteristics of subordinate-level recognition that undergo pronounced change with practice are illustrated schematically in figures 1 and 2. These are the phenomena of canonical views (Palmer et al., 1981; Edelman et al., 1989) and pseudo mental rotation (analogous to the "classical" mental rotation of (Shepard and Metzler, 1971); see (Tarr and Pinker, 1989; Edelman and Weinshall, 1991)).

1.2.1 Canonical views

Three-dimensional objects are more easily recognized when seen from certain viewpoints, called canonical, than from other, random, viewpoints (Figure 1). The advantage of canonical views is manifested in consistently shorter response time, lower error rate and higher subjective "goodness" rating (Palmer et al., 1981). Canonical views

[1] We have observed improvement in error rates in the absence of feedback, in situations where relatively small perceptual "distance" between clearly recognizable and difficult views allowed the subject to consciously deduce the identity of a difficult view. Once such a view became easily recognizable, it could be used to facilitate recognition of yet another difficult view in a kind of "stepping-stone" strategy (Edelman and Bülthoff, 1990).

Figure 1: Canonical views: certain views of 3D objects are consistently easier to recognize or process in a variety of visual tasks (Palmer et al., 1981). For example, a front view of a pair of spectacles is expected to yield lower response time and error rate and to receive higher subjective "goodness" score than a top view of the same object. Such differences may exist even among views that are seen equally often.

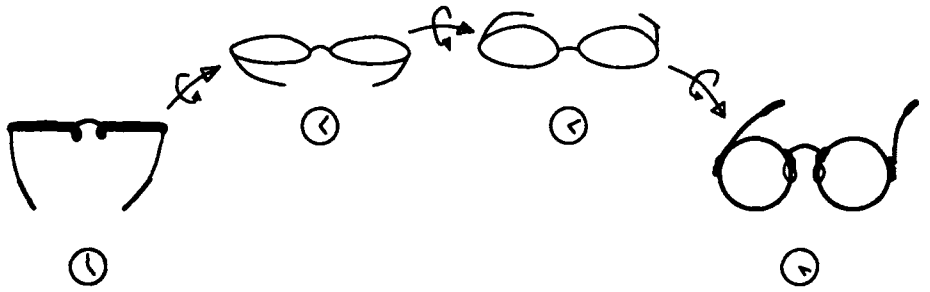

Figure 2: Recognition time for an object grows monotonically with its misorientation relative to a canonical view, as if the object is mentally rotated to match an internal representation. Rates of "rotation" range between $40°/sec$ and $550°/sec$ (see (Tarr and Pinker, 1989)), depending on the stimuli and the task. This effect tends, however, to disappear with practice (Tarr and Pinker, 1989; Edelman et al., 1989).

are routinely found for synthetic novel objects under controlled exposure conditions, even when each view is shown equally often (Edelman et al., 1989).

1.2.2 Pseudo mental rotation

Transition from a canonical to a non-canonical view of an object does not merely increase the expected recognition time. Rather, response latency depends on the viewpoint in an orderly fashion, growing monotonically with misorientation relative to the nearest canonical view (Figure 2; for a review see (Tarr and Pinker, 1989; Edelman and Bülthoff, 1990)). This dependency of response time on misorientation resembles the finding by Shepard and Metzler (Shepard and Metzler, 1971) of a class of phenomena that became known as mental rotation (see (Shepard and Cooper, 1982) for an

overview).[2]

1.3 Effects of practice on canonical views and mental rotation

While uniform initial exposure does not preclude the formation of canonical views, repeated presentation of the same stimulus eventually increases the uniformity of response time over different views of the stimulus. Thus, practice affects the response time aspect of the canonical views phenomenon: after only a few trials, the differences in response time between canonical and random views diminish significantly, even in the absence of any feedback to the subject (Edelman et al., 1989). Notably, the differences in error rate remain fairly constant.

Mental rotation in recognition is also affected by practice. For example, Tarr and Pinker (1989) found that repeated exposure to the same stimulus caused an apparent shift in the subject's strategy: while naming time for novel test views grew monotonically with misorientation relative to the nearest training view, familiar test views yielded essentially constant response times (this is consistent with a changeover from time-consuming rotation-based strategy to a faster memory-intensive approach that saves time by storing all frequently occuring views). A similar effect has been reported by Edelman et al.(Edelman et al., 1989; Edelman and Weinshall, 1991). It has also been shown (Edelman and Weinshall, 1991) that both the initial manifestation of mental rotation and its disappearance with exposure can be replicated by a model that does not rely on 3D object-centered representations and, a fortiori, has no means for rotating such representations.

1.4 Scope of the present report

Should the weakening of canonical views with practice be attributed to the formation or the modification of object-specific representations in the visual system (as several models of recognition would have it (Tarr and Pinker, 1989; Edelman and Weinshall, 1991)), or is it merely a manifestation of increased task-related proficiency? Previous investigations (Jolicoeur, 1985; Larsen, 1985; Tarr and Pinker, 1989) have found that the practice-related change in the dependency of response time on object orientation was object-specific. However, response times in a variety of psychophysical tasks are affected also by factors that are not stimulus-specific (Luce, 1986). The aim of the present study was to separate task (or proficiency) effects from object-specific effects on mean response time and several performance measures derived from it. For that purpose, we inspected the development of mean response time, its dependency on object orientation, and its variation over object views, across four blocks of trials. In these

[2] In Shepard's experiments the task was to determine whether two simultaneously shown images were projections of the same 3D object, or of different objects related by a mirror transformation. In this respect, classical mental rotation is different from recognition experiments, where only a single object is presented at a time.

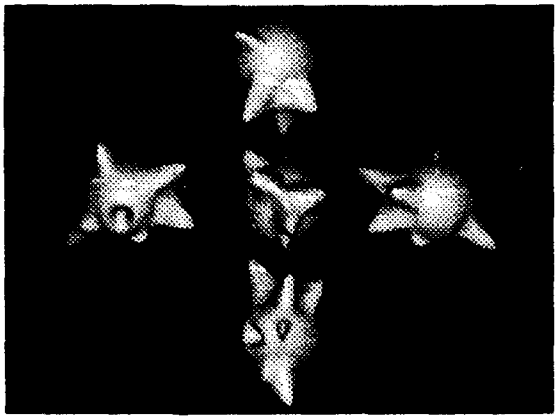

Figure 3: An example of the amoeba-like stimuli used in the experiment. These 3D shapes were created by a computer graphics program which generated between four and eight protrusions or indentations on a sphere. The image in the center represents one view of a amoeba-like 3D object. The other images are derived from the same object by $\pm 75°$ rotation around the vertical or horizontal axis. The objects were rendered during testing, in real time, on a graphics monitor.

four blocks, two different object sets (say, 1 and 2) were shown in an alternating fashion (that is, blocks $1, 2, 3, 4$ corresponded to object sets $1, 2, 1, 2$). In this arrangement, the change in performance from block 1 to block 2, and from 3 to 4 could be interpreted as task learning, while the change between the means of blocks $1, 2$ and $3, 4$ could be due both to object-specific and to task learning. Thus, if a measure of the strength of canonical views turns out to be unaffected by the transition between blocks 1 and 2, but changes between blocks 1,2 and 3,4, it would reflect more closely changes in object representation rather than changes in general recognition proficiency.

2 Experimental method

2.1 Subjects

Five subjects participated in this experiment (see Table 1). Two of the subjects (smd and cda) were inexperienced observers, and were naive as to the issues under study. The other three had had varying amounts of previous experience in visual psychophysics.

2.2 Stimuli

The stimuli were computer graphics renderings of irregular three-dimensional objects ("amoebae"; see Figure 3). Forty-eight amoebae were generated by creating between four and eight random extrusions and/or indentations on a standard sphere. This

Session →	*Session 1*		*Session 2*	
Block of 3 trials per object →	*block 1*	*block 2*	*block 3*	*block 4*
Order group 1 (subjects dls, isa, smd)	set **1**	set **2**	set **1**	set **2**
Order group 2 (subjects cda, gjz)	set **2**	set **1**	set **2**	set **1**

Table 1: Summary of the combinations of factors tested in the experiment.

process was carried out on a Symbolics Lisp Machine in the S-Geometry programming environment. The geometrical descriptions of the objects were then transferred to a different machine for rendering.

2.3 Apparatus

A Stellar GS-1000 graphics computer was used to present the stimuli and to record responses. The Stellar's graphics monitor utilized a P22 phosphor and ran at a refresh rate of 74 Hz (non-interlaced). Upon presentation, the base sphere subtended $9cm$ on the monitor screen. Maximally extruded amoebae could span an additional $9cm$. Subjects viewed the stimuli from a distance of approximately $65cm$ and were unrestrained. Reaction times were derived from keyboard button presses using the XEvent facilities of XWindows.

2.4 Design

Two test sets were constructed from the same set of 48 stimuli. Each set utilized a distinct collection of six target objects, and a common pool of 36 distractor objects. Subjects completed two sessions, each containing both test sets. The sessions took place $1-3$ days apart. The order of presentation of the two test sets in each session was varied between subjects. Table 1 summarizes the experimental design.

2.5 Procedure

In each session there were two blocks of test trials, one per test set. Each block consisted of six repetitions (one per target object) of a training phase, followed by a test phase. Within each sequence, the training phase was signaled by the appearance of a verbal prompt ("Training Phase n", where n was the target number). The subject initiated the training phase by pressing a key on the keyboard. During training, the target object was imaged at the center of the display. The target object was shown rotating about its vertical axis for three full oscillations of $\pm 90°$ around an arbitrary fixed reference orientation. The rate of rotation was approximately $20°/sec$. The total exposure time of each target object during its training phase was approximately $30sec$.

The test phase of each sequence was also initiated by the subject, upon the presentation of a verbal prompt. Each of the 11 possible views ($-90°$ to $+90°$, at $18°$ intervals) of the target object for that sequence was then shown three times, in a random order. The 33 target views were randomly intermixed with 33 random views belonging to the six distractors assigned to that target. The subject's task was to decide whether the displayed view belonged to the target shown previously in the training phase, or not (two-alternative forced choice). Subjects indicated their choice by pressing one of two buttons on a standard computer keyboard. Response caused the disappearance of the test image, and the appearance of a central fixation cross. The fixation cross was present for a $500ms$ interval, and was followed by the next test image.

3 Experimental results

3.1 Error rate effects

Since all the test views in the experiment have been shown to the subjects during training, generalization (Bülthoff and Edelman, 1992) was not an issue. Consequently, the distribution of error rates for different views is of secondary importance, and its detailed description has been relegated to the appendix. We mention at this point only that the mean error rate was 7.8%, and that there was no evidence of speed-accuracy trade-off, as the response times and the error rates exhibited similar dependence on misorientation relative to the canonical view.

3.2 Raw response time effects

Response time (RT) was the main dependent variable of interest in the experiment. RT was measured from the moment the stimulus appeared on the screen to the subject's key press. RTs shorter than $250ms$ or longer than $3000ms$ were discarded. Mean RT ranged from $874 \pm 15ms$ in session 1 to $553 \pm 15ms$ in session 2. Overall mean response time in the experiment was $713ms$. Note that these short RTs are a good indication that the subjects' response was automatic (unlike in the classical mental rotation experiments (Shepard and Metzler, 1971)).

3.2.1 Development of response time with practice

As expected, RT exhibited continuous decrease throughout the four experimental blocks, both for order 1 and 2 (see Figure 4). Clearly, subjects became more proficient in solving the given recognition task, since the RTs decreased from each block to the next, without regard to the identity of the target set in each block. The decrease in RT and its eventual flattening-out in session 2 signify the extent of task learning that

happened in the experiment.[3] Notably, the asymptotic value of RT was the same for both groups of subjects, even though their RTs in block 1 differed significantly.

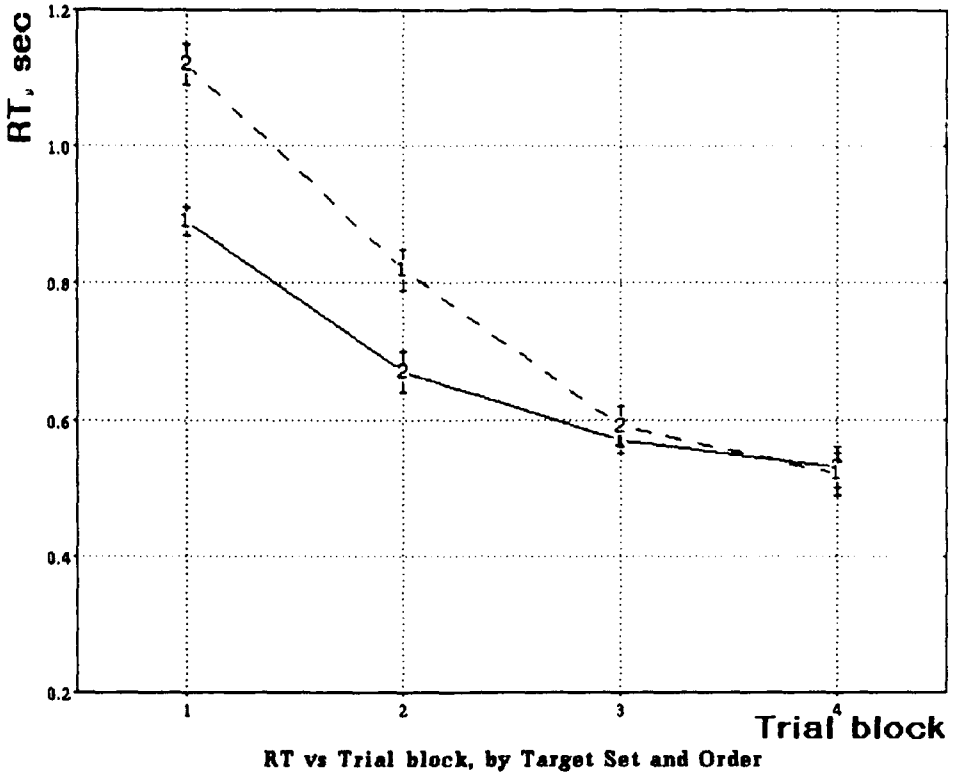

Figure 4: Response time (RT) in the four blocks of the experiment, plotted separately for order 1212 (solid line) and 2121 (dashed line). Recall that different subject groups were tested in these two conditions (see Table 1). Error bars here and below show ± 1 standard error of the mean.

3.2.2 Pseudo mental rotation effect in RT and its development with practice

As mentioned in section 1, it has been previously found that RT for a given view of a 3D object increases monotonically with its misorientation relative to the closest familiar or canonical view (Tarr and Pinker, 1989). Furthermore, this dependence tends to disappear with practice after only a few exposures of the subject to the testing views.

[3]To demonstrate this aspect of task learning more forcefully, it would be useful to test the progress of RT in a prolonged experiment in which a new set of objects would be presented in each successive session.

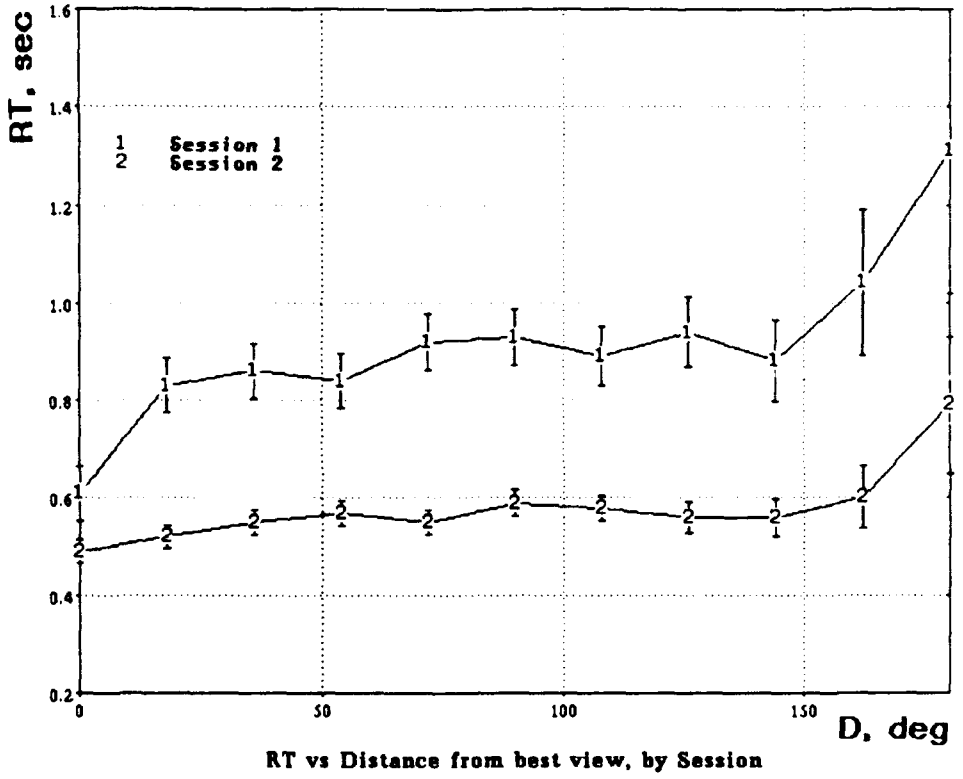

Figure 5: The dependence of response time (RT) on misorientation D relative to a canonical view (see text), by session. Regression of RT on D showed that the significance of "mental rotation" was greatly reduced in session 2, compared to session 1.

In three dimensions, this *pseudo mental rotation* phenomenon has been demonstrated only for stick or wire-like objects (Tarr and Pinker, 1989; Edelman et al., 1989). We have reproduced both the existence of pseudo mental rotation and its weakening with practice with the amoeba-like targets of the present experiment.

The analysis of the dependence of RT on object orientation proceeded as follows. First, a (subject-specific) canonical view was identified for each target object as the test view that yielded the shortest RT in session 1. Next, every view was assigned a *distance* D from the best-RT view, defined as the absolute value of the angular misorientation between the two. Finally, RT was plotted against D, separately for sessions 1 and 2 (see Figure 5). The dependence of RT on D was assessed by computing the regression coefficients of D (see appendix B for a discussion of this method). The approximation for session 1 resulted in a reciprocal rotation rate (the coefficient of D)

of $4.2 \pm 1.4ms/°$ ($T(1,465) = 2.9, p < 0.004$).[4] In comparison, in session 2 both the reciprocal of rotation rate and its significance were greatly reduced: the coefficient of D was $1.3 \pm 0.6ms/°$, $T(1,470) = 2.1, p < 0.04$. We remark that in an analogous experiment with wire ..ke objects five to ten trials per view were necessary to obliterate completely the effect of mental rotation in recognition (Edelman et al., 1989).

3.3 Normalized response time effects

The main statistical measure of performance that allowed us to compare task and object learning in this experiment was the *coefficient of variation* of response time over views of the target (CVRT), defined as the standard deviation of RT divided by its mean. Previous studies of the development of recognition with practice (Edelman et al., 1989; Edelman and Weinshall, 1991) employed the CVRT measure as an indicator of the strength of the canonical views phenomenon, claiming that it captures the variability of response time over different views of an object, while discounting the effect of general proficiency (as seen in the decrease of mean response time). The present analysis attempted to assess the degree to which the CVRT measure reflects the structure of object representations, by separating object-specific decrease in CVRT from its non-specific component (presumably, related to task proficiency).

The effects of practice on CVRT were assessed by a General Linear Models (GLM) procedure, in a 3-way (Session × Target-set × Order) repeated measures analysis of variance. According to the experimental design described in the previous section, Session was a within-subjects effect, while the effects of Target-set and Order were between-subjects.

The two sought-after types of practice effects — object learning and task learning — can be distinguished in the experimental data as follows. First, a decrease in CVRT from target set 1 to 2 in the first row of Table 1, and a similar decrease from set 2 to 1 in the second row of the table), would signify task learning, because that would mean that being in the first or in the second block in a given session is a stronger determinant of CVRT than random factors having to do with object identity (see Figure 6). Second, if no task effect on CVRT is found, object-specific learning would be apparent in a significant main effect of Session. Finally, as a control against the possibility that the learning effects are an artifact of presentation order, the *lack* of Order × Target-set interaction should be established.

As Figure 6 shows, the experimental data indicate that CVRT is indeed an appropriate measure of the strength of canonical views (see also appendix C), since it appears to be susceptible to object-specific learning effects, but not to general proficiency of the subject (task learning). The analysis of variance supports this conclusion: the only significant effects for CVRT it reports are those of Session ($F(1,56) = 6.27; p < 0.02$)

[4]The coefficient of D^2 was also computed, and found to be very small, negative and of marginal significance ($p = 0.08$).

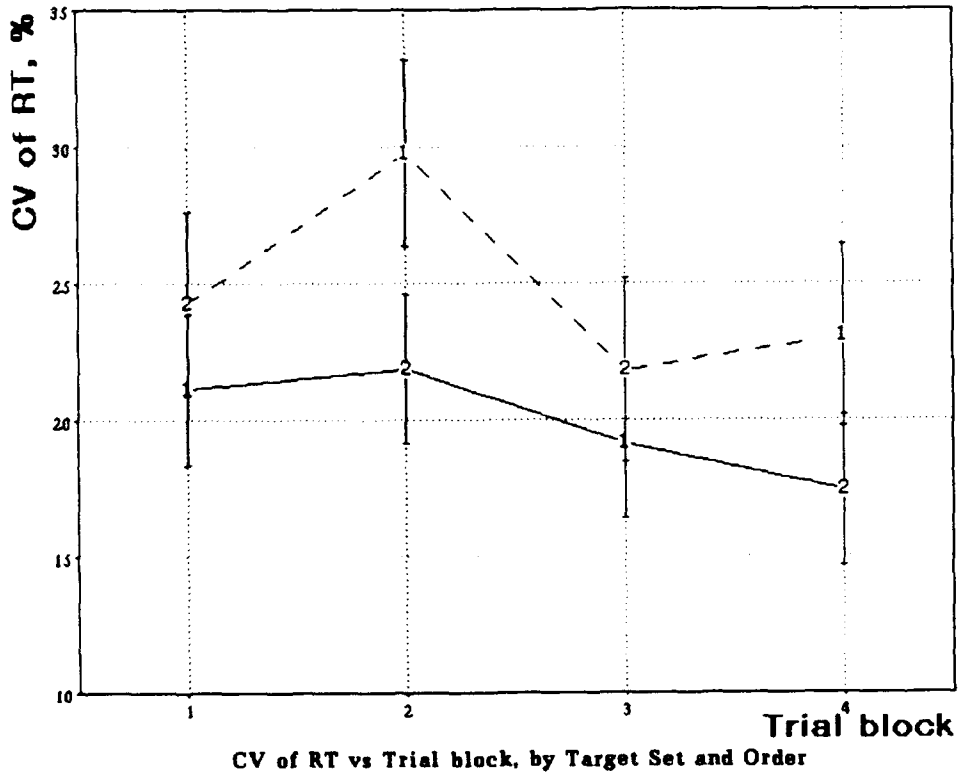

Figure 6: The logic behind the interpretation of the experimental results described in section 3 is illustrated in this figure, which shows a plot of the development of CVRT for the two groups of subjects. The first group (solid line) was shown target sets in the order 1, 2, 1, 2, while the order for the second group (dashed line) was 2, 1, 2, 1 (see Table 1). Although the mean CVRT for the two groups was different, the effects of learning were the same. Specifically, the absence of decrease in CVRT between blocks 1 and 2 and between blocks 3 and 4 indicates that the CVRT measure is unaffected by task learning. At the same time, the decrease in CVRT between the average of blocks 1 and 2 on one hand, and blocks 3 and 4 on the other hand (corresponding to the significant Session effect in the analysis of variance in section 3) indicates pronounced object learning. In other words, CVRT decreases (that is, the canonical views phenomenon weakens) only if *the same object set* is seen for a second time.

and of Order ($F(1,56) = 3.36; p < 0.07$). As we have mentioned, the Session effect represents object learning, while the Order effect is irrelevant to the issue of learning (see also the caption to Figure 6) and has to do with between-subjects differences.

4 Summary

The goal of the present study was to quantify and and compare effects of task and object learning in visual recognition. Experimental results we have discussed indicate that these two types of learning may be treated separately, provided that appropriate measures are employed for their quantification. In particular, the effect of task learning, as measured by the reduction in the mean response time, appears to saturate after a few trials, when the response time reaches its asymptotic value (about $550ms$ in our experiments). Importantly, mean RT is reduced no matter what the target objects are (that is, mean RT as a measure of task learning is object-nonspecific, as expected). On the other hand, the development of the coefficient of variation of RT over views with practice, which is a useful estimate of the development of canonical views, appears to be indeed an object-specific measure of learning.

An immediate benefit of this finding is a corroboration of our earlier results regarding the development of canonical views with practice (Edelman et al., 1989), this time with amoeba-like rather than wire-like objects. Specifically, practice-induced modification of object representations (Edelman and Weinshall, 1991) gains thereby plausibility as an explanation of the weakening of canonical views and of the pseudo mental rotation phenomena associated with them.

Acknowledgement

We thank Shimon Ullman for his extensive comments on a draft of this report.

Appendix A: Error rate effects

As we have explained above, error rate effects are of secondary relevance in the present experiment, in which only previously seen views were shown, and in which the subjects received no feedback as to the correctness of their responses. Consequently, the only two characteristics of the pattern of error rates that we mention here are the absence of speed-accuracy trade-off, important for the assessment of the validity of response time data, and the independence of error rate on practice.

Absence of speed-accuracy trade-off

Evidence in favor of the conclusion that shorter response time was not traded off for higher error rate can be seen in Figure 7, which shows the dependence of ER on misorientation D relative to the *shortest-RT* view (compare with Figure 7). Clearly, the best-RT view is also the best as far as ER is concerned. Furthermore, neither this, nor the dependence of ER on D was affected by practice.

Independence of error rate on practice

The mean error rate in the experiment was not affected by practice (see Figure 8), which is expected given that there was no feedback to the subjects. Neither did practice have any significant effect on the coefficient of variation of ER over views (as found in a similar experiment with wire-like objects (Edelman et al., 1989)).

Appendix B: quantifying the dependence of recognition on orientation using regression analysis

If the canonical views of an object are known, the dependence of, say, response time on misorientation relative to those views can be assessed by regression analysis. In some cases, however, the canonical views are unknown in advance. For example, given the strong influence of practice on the pattern of RT for a novel object, one may wish to minimize the exposure of the subject to the stimuli prior to the experiment. In this situation, it is still possible to use regression to characterize the dependence of RT on orientation, provided that special care is taken in interpreting the results. The analysis then proceeds as follows. First, a canonical view is defined operationally as the view that yields the shortest RT. Second, every test view is assigned a number D equal to its distance (misorientation) from the canonical view. Third, the regression of RT on D (and, if necessary, on D^2) is computed (see Figure 9). The statistical significance of the regression is assessed using an analysis of variance. The F-ratio statistic for the null hypothesis that all the regression coefficients except the intercept are equal to 0 is given by the expression (Mendenhall and Sincich, 1988):

$$F = \frac{R^2/k}{(1-R^2)/[n-(k+1)]} \tag{1}$$

where n is the number of data points, $k+1$ is the order of regression, and R^2 is the coefficient of determination of the regression, defined as

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \tag{2}$$

with y_i, $\hat{y}_i$, and $\bar{y}$ being the true, predicted and average values of the dependent variable.

If the regression, as characterized by the F-ratio computed using eq. (1), turns out to be significant, it means that the orderly dependence of RT on D is unlikely to be due to an accidental pattern or artifact in RT data. Rejecting the null hypothesis using the above F-test, then, provides a solid ground for the claim that at least one additive component of RT has a near linear dependence on D.[5] Two further controls regarding this interpretation of the regression results can be made as follows (Edelman et al., 1989). First, the regression fails if a view other than the shortest-RT one is chosen as the reference view. Second, having a large variation in RT over views is in itself insufficient to ensure that the RT will depend smoothly on the distance to the shortest-RT view. This last conclusion is supported by the outcome of the regression following practice (e.g., in the second session in our experiments).

[5]The significance of the regression does not necessarily indicate that this dependence stems from the "real" mental rotation, reported, e.g., by Cooper (1976) using the intermediate probe method. It does indicate, however, the statistical reliability of this dependence.

Appendix C: quantifying the strength of the canonical views effect using the CV measure

The existence of canonical views for objects that are seen equally often from all the attitudes at which their recognition is tested (Edelman and Bülthoff, 1991) implies that some views of those objects are inherently easier to recognize than others. To quantify this difference among views, we use the *coefficient of variation* (CV), defined as the standard deviation divided by the mean, of the response time (or of the error rate). The CV measure for a given object is computed in two steps. First, the means m_{v_i} for each of the n views are computed separately. Second, the CV is obtained as the ratio of the standard deviation of m_{v_i} to the grand mean $M = \frac{1}{n}\sum_{i=1}^{n}(m_{v_i})$, both taken over all the different views under consideration:

$$CV = \frac{std(m_{v_i})}{M} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(m_{v_i} - M)^2}}{M} \tag{3}$$

The standard deviation $std(m_{v_i})$, which is computed over the different views, should not be confused with standard deviation with respect to other independent variables in an experiment, such as time (for which the standard deviation is usually found to covary with the mean; see, e.g., Luce, 1986, p.64). By its definition, the value of $std(m_{v_i})$ captures the intuitive notion of variability of responses over views. Division by the grand mean, as done in eq. (3), allows the comparison of this measure of variability over views *across objects*, each of which may have a different mean, reflecting the general difficulty of its recognition.

References

Biederman, I. (1987). Recognition by components: a theory of human image understanding. *Psychol. Review*, 94:115–147.

Bülthoff, H. H. and Edelman, S. (1992). Psychophysical support for a 2-D view interpolation theory of object recognition. *Proceedings of the National Academy of Science*, 89:60–64.

Cooper, L. (1976). Demonstration of a mental analog of an external rotation. *Perception and Psychophysics*, 19:296–302.

Edelman, S., Bülthoff, H., and Weinshall, D. (July 1989). Stimulus familiarity determines recognition strategy for novel 3D objects. A.I. Memo No. 1138, Artificial Intelligence Laboratory, Massachusetts Institute of Technology.

Edelman, S. and Bülthoff, H. H. (1990). Viewpoint-specific representations in 3D object recognition. A.I. Memo No. 1239, Artificial Intelligence Laboratory, Massachusetts Institute of Technology.

Edelman, S. and Bülthoff, H. H. (1991). Orientation dependence in the recognition of familiar and novel views of 3D objects. *Vision Research*. submitted.

Edelman, S. and Weinshall, D. (1991). A self-organizing multiple-view representation of 3D objects. *Biological Cybernetics*, 64:209–219.

Jolicoeur, P. (1985). The time to name disoriented objects. *Memory and Cognition*, 13:289–303.

Larsen, A. (1985). Pattern matching: effects of size ratio, angular difference in orientation and familiarity. *Perception and Psychophysics*, 38:63–68.

Luce, R. D. (1986). *Response times: their role in inferring elementary mental organization*. Oxford University Press, Oxford.

Mendenhall, W. and Sincich, T. (1988). *Statistics for the engineering and computer sciences*. Macmillan, London.

Palmer, S. E., Rosch, E., and Chase, P. (1981). Canonical perspective and the perception of objects. In Long, J. and Baddeley, A., editors, *Attention and Performance IX*, pages 135–151. Erlbaum, Hillsdale, NJ.

Rock, I. and DiVita, J. (1987). A case of viewer-centered object perception. *Cognitive Psychology*, 19:280–293.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., and Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8:382–439.

Shepard, R. N. and Cooper, L. A. (1982). *Mental images and their transformations.* MIT Press, Cambridge, MA.

Shepard, R. N. and Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171:701–703.

Tarr, M. and Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, 21:233–282.

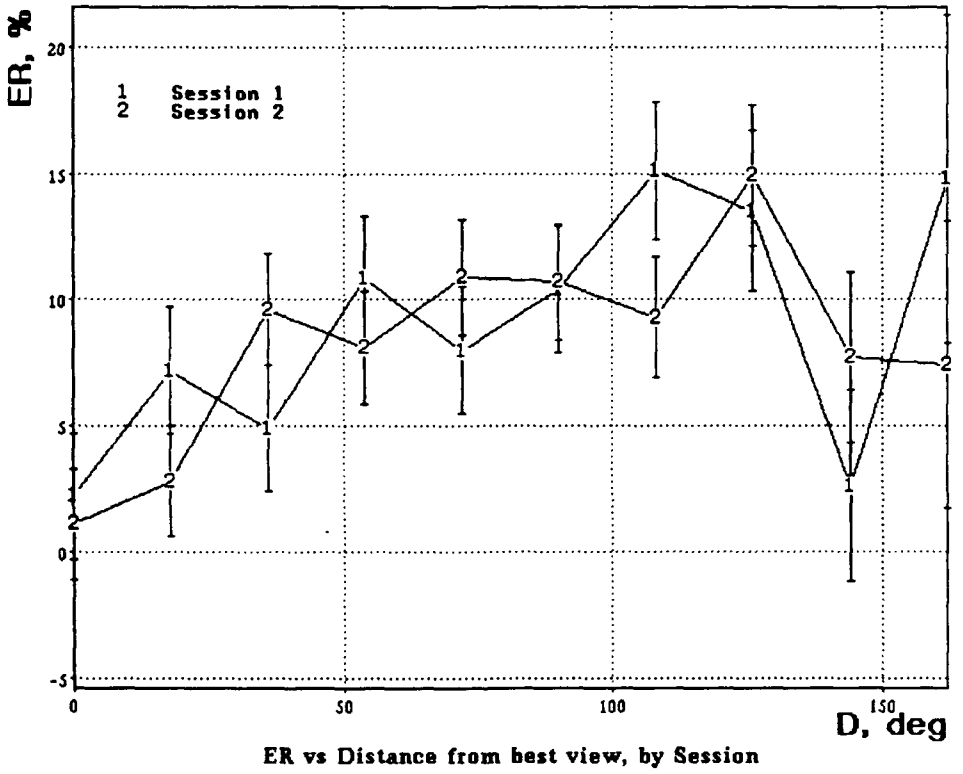

Figure 7: The dependence of error rate (ER) on the distance D to the best (shortest-RT) view, plotted by session.

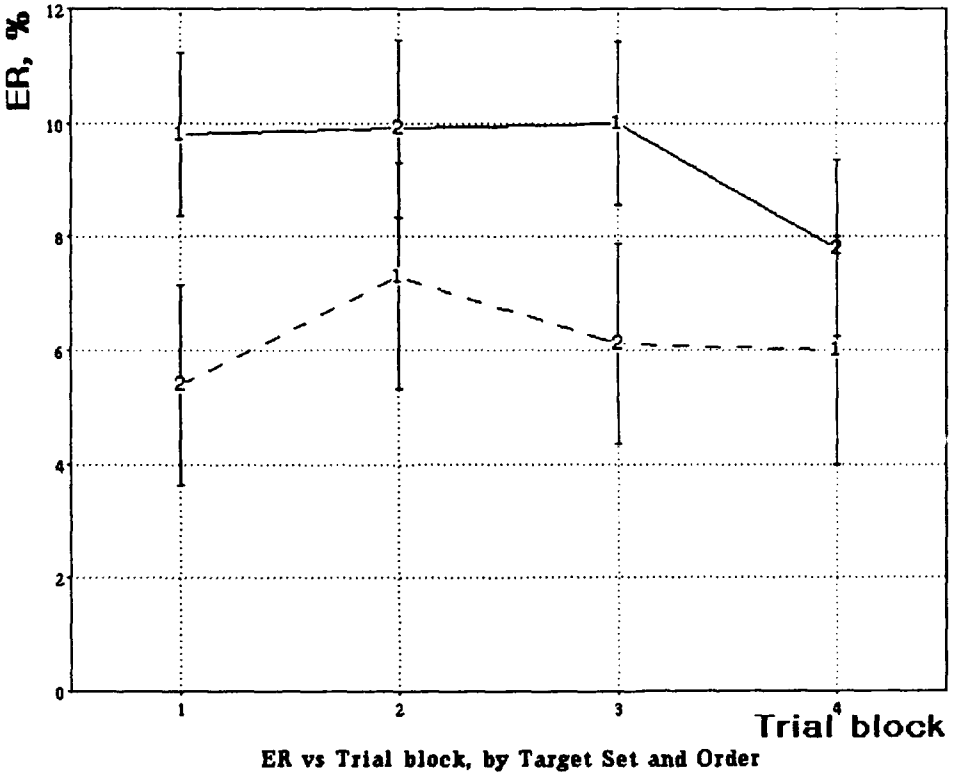

Figure 8: The development of error rate (ER) with practice. Blocks 1 and 2 correspond to session 1; blocks 3 and 4 — to session 2. There is no effect of block, as expected in the absence of feedback to the subjects.

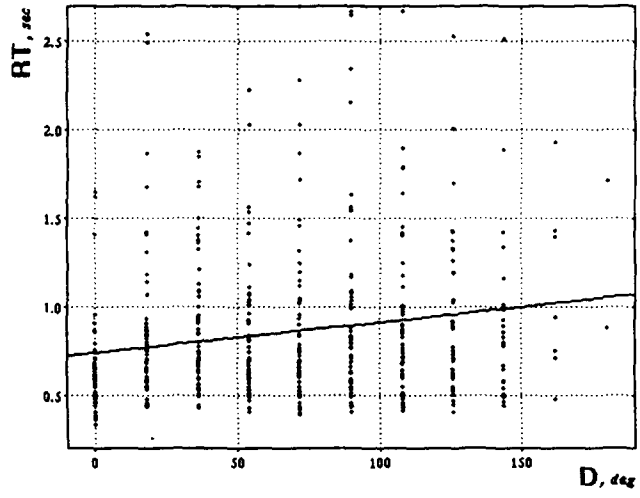

Figure 9: Scatter plot of RT vs. D for session 1, and the best linear fit to the data. The significance of this regression is $p < 0.0001$ ($F(1, 466) = 15.9$). Thus, despite the low value of the coefficient of determination ($R^2 = 0.033$), the linear trend in the data is highly reliable, due to the large number of data points ($n = 468$).